Exchangeability and continuum limits of discrete random structures

David Aldous

August 22, 2010

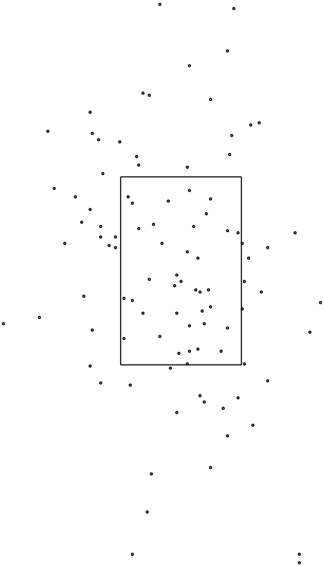
- Exchangeability and de Finetti's theorem (1930s - 50s)
- Structure theory for exchangeable arrays (1980s)
- A general program for continuum limits of discrete random structures, illustrated by trees (1990s)
- 3 novel "pure math" developments (2000s)
- Google map routes from this viewpoint (2010s)

1. Exchangeability and de Finetti's theorem

Theorem (de Finetti)

Each infinite exchangeable sequence of RVs is distributed as a mixture of IID sequences

For those who don't work in Probability let me try to explain what this means.



99 dart throws, centered on a 2.25" × 3.5" playing card.

Key point (for this talk): The following two scenarios are different.

- Pick random person in audience; ask to throw dart 10 times.
- Pick random person in audience; ask to throw dart 1 time. Repeat 10 times (so probably 10 different people).

Model for dart throws:

- For each person there is a probability measure μ on $\mathbb{R}^2$; the chance their dart lands in A equals $\mu(A)$.
- When this person throws repeatedly, the landing points $X_1, X_2, \ldots$ are independent random variables with distribution μ.

Independence formalized by *product rule*

$$P(X_1 \in A_1 \text{ and } X_2 \in A_2) = \mu(A_1) \times \mu(A_2)$$

or equivalently *product measure*

$$\text{dist}(X_1, X_2) = \mu \otimes \mu \qquad \text{dist}(X_1, X_2, X_3, \ldots) = \mu^{\otimes \infty}.$$

Back to 3 scenarios.

- Pick person with known μ; ask to throw dart for ever.
- Pick random person in audience; ask to throw dart 1 time. Repeat for ever.
- Pick random person in audience; ask that person to throw dart for ever.

The three scenarios give three different distributions for the infinite sequence $(X_1, X_2, \ldots)$. With a 500-person audience

- $\mu^{\otimes\infty}$
- $\nu^{\otimes\infty}$ where $\nu(\cdot) = \frac{1}{500}\sum_k \mu_k(\cdot)$
- $\frac{1}{500}\sum_k \mu_k^{\otimes\infty}$

In the first two cases, different throws are independent, but in the third they're not. Jargon: in first two cases the distribution is IID (independent and identically distributed), third case is *mixture of IID*.

The third case is central to this talk, but first more conceptual and mathematical background.

Conceptual points.

View a probability measure (PM) as like a recipe or a plan – something you *might* do – and a random variable (RV) as an instance of actually doing it – in freshman statistics we say "chance experiment".

RVs can take values in an (essentially) arbitrary space S.
A S-valued RV X has a *distribution* $\text{dist}(X)$, the induced PM on S.

Most definitions in Probability Theory are formally about PMs but we apply them to RVs.

When we talk about symmetry properties we are talking about an underlying PM not the realizations of RVs.

Imagine idealized random number generator (RNG) that gives a random number ξ distributed uniformly on $[0,1]$; repeated calls to the RNG give independent $\xi_1, \xi_2, \ldots$.

Given an arbitrary (measurable) function $f : [0,1] \to S$ for a "nice" space S, we can use $f(\xi)$ as a S-valued RV with some distribution μ. Different f might give the same μ.

An under-emphasized Theorem in measure theory says that every μ arises as $\mathrm{dist}(f(\xi))$ for some f.

Any time we do a computer simulation of a probability model we are implicitly using the latter fact!

Summary so far: an IID S-valued sequence arises as $(f_1(\xi_1), f_1(\xi_2), \ldots)$ where the $(\xi_1, \xi_2, \ldots)$ – view as calls to a RNG – are IID uniform$[0,1]$, and where $f_1 : [0,1] \to S$ is some function.

Definition of "mixture of IID sequences" is a PM on S^∞ of form

$$\int \mu^{\otimes\infty} \ \Psi(d\mu), \ \text{some PM } \Psi \text{ on } \mathcal{P}(S) := \{\text{PMs on } S\}$$

and as Corollary, any such PM has a representation as the distribution of

$$f_2(\alpha, \xi_1), \ f_2(\alpha, \xi_2), \ f_2(\alpha, \xi_3), \ldots \quad (1)$$

for some function $f_2 : [0,1] \times [0,1] \to S$. Here α is one more independent uniform $[0,1]$ RV.

(This relies on "measure theory magic" – the action of picking a PM at random is implemented as a function of α).

A finite permutation π of $\{1, 2, 3, \ldots\}$ induces a map $\tilde{\pi} : S^\infty \to S^\infty$, mapping (s_i) to $(s_{\pi(i)})$.

Definition

A PM on S^∞ is **exchangeable** if it is invariant under the action of each $\tilde{\pi}$.

Intuitively "order of RVs does not matter". This is a strong symmetry condition.
Obvious that any (finite or infinite length) mixture of IID sequences is exchangeable.

Theorem (de Finetti)

Each infinite exchangeable sequence is distributed as a mixture of IID sequences

and so has a **representation** in form (1)

$$f_2(\alpha, \xi_1),\ f_2(\alpha, \xi_2),\ f_2(\alpha, \xi_3), \ldots \qquad (1)$$

de Finetti's Theorem appears in some first-year-grad probability textbooks, and plays a conceptually fundamental role in Bayesian Statistics, which I won't explain in this talk.

The next material is somewhat deeper, and goes back to Hoover (1979, unpublished) and Aldous (1981), then developed more systematically by Kallenberg (1989+) culminating in his 2005 monograph *Probabilistic Symmetries and Invariance Principles*.

The motivation was in part "mathematically natural conjectures", in part Bayesian statistics. I won't explain the original motivation in this talk, because more recent uses are more interesting.

2. Structure theory for partially exchangeable arrays

Write $\mathbb{N} := \{1, 2, 3, \ldots\}$ and write $\mathbb{N}_{(2)}$ for the set of unordered pairs $\{i, j\} \subset \mathbb{N}$. Consider a random array

$$\mathbf{X} = (X_{\{i,j\}},\ \{i, j\} \in \mathbb{N}_{(2)}).$$

(Essentially a random infinite symmetric matrix). We want to study the the *partially exchangeable* property

$$\mathbf{X} \stackrel{d}{=} (X_{\{\pi(i),\pi(j)\}},\ \{i, j\} \in \mathbb{N}_{(2)}) \text{ for each finite permutation } \pi. \quad (2)$$

Because not every permutation of $\mathbb{N}_{(2)}$ is of the form $\{i, j\} \to \{\pi(i), \pi(j)\}$, this is a weaker property than exchangeability of the countable family $\mathbf{X}$.

$\mathbf{X} \stackrel{d}{=} (X_{\{\pi(i),\pi(j)\}}, \ \{i,j\} \in \mathbb{N}_{(2)})$ for each permutation π

We can create such a partially exchangeable array by starting with our IID uniform[0, 1] RVs $(\xi_1, \xi_2, \ldots)$ and applying a function $g_2 : [0,1]^2 \to \mathbb{R}$ which is symmetric in the sense $g_2(x,y) = g_2(y,x)$, to get

$$X_{\{i,j\}} = g_2(\xi_i, \xi_j).$$

This is the "interesting" construction of an array with the partially exchangeable property. But also there are the arrays

- with IID entries
- where all entries are the same RV.

We can combine these ideas as follows.
Take a function $f : [0,1]^4 \to S$ such that $f(u, u_1, u_2, u_{12})$ is symmetric in (u_1, u_2), and then define

$$X_{\{i,j\}} := f(U, U_i, U_j, U_{\{i,j\}}) \tag{3}$$

where all the r.v.'s in the families $U, (U_i, i \in \mathbb{N}), (U_{\{i,j\}}, \ \{i,j\} \in \mathbb{N}_{(2)})$ are IID Uniform(0, 1). The array $\mathbf{X} = (X_{\{i,j\}})$ is partially exchangeable.

Structure theory developed (1980s) by Hoover - Aldous - Kallenberg covers many variants of the following result, detailed in Kallenberg (2005).

Theorem (Partially Exchangeable Representation Theorem)

An array $\mathbf{X}$ *which is partially exchangeable, in the sense (2), has a representation in the form (3).*

$$\mathbf{X} \stackrel{d}{=} (X_{\{\pi(i),\pi(j)\}}, \ \{i,j\} \in \mathbb{N}_{(2)}) \text{ for each finite permutation } \pi. \qquad (2)$$

$$X_{\{i,j\}} := f(U, U_i, U_j, U_{\{i,j\}}) \qquad (3).$$

There is a (technically complicated) uniqueness property - f is unique up to measure-preserving transformations of the U's.

3. Continuum limits of discrete random structures

This is a "general program", where "general" $\neq$ "always works" but instead means "works in various settings that otherwise look different".

Rather obvious idea:

> *One way of examining a complex mathematical structure with a PM is to sample IID random points and look at some form of induced substructure relating the random points*

which assumes we are **given** the complex structure.

Less obvious idea:

> *We can often use exchangeability in the* **construction** *of complex random structures as the $n \to \infty$ limits of random finite n-element structures $\mathcal{G}(n)$.*

What's the point? Use when there's no natural way to think of each $\mathcal{G}(n)$, as n varies, as taking values in the **same** space.

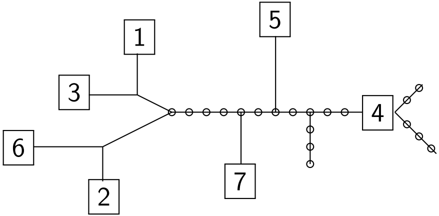
To expand the idea:

> *Within the n-element structure $\mathcal{G}(n)$ pick k IID random elements, look at an induced substructure on these k elements – call this $\mathcal{S}(n,k)$ – taking values in some space $S_{(k)}$ that depends on k but not n. Take a limit (in distribution) as $n \to \infty$ for fixed k, any necessary rescaling having been already done in the definition of $\mathcal{S}(n,k)$ – call this limit $\mathcal{S}_k$. Within the limit random structures $(\mathcal{S}_k, 2 \le k < \infty)$, the k elements are exchangeable, and the distributions are consistent as k increases and therefore can be used to* **define** *an infinite structure $\mathcal{S}_\infty$.*

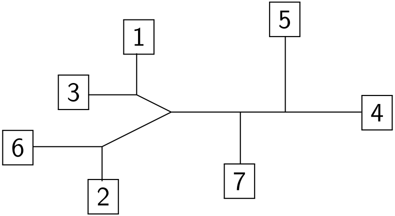
Where one can implement this program, the random structure $\mathcal{S}_\infty$ will for many purposes serve as a $n \to \infty$ limit of the original n-element structures. Note that $\mathcal{S}_\infty$ makes sense as a rather abstract object, via the Kolmogorov extension theorem, but in concrete cases one tries

- to identify $\mathcal{S}_\infty$ with some more concrete construction
- to characterize all possible limits of a given class of finite structures.

Trees fit nicely into the "substructure" framework. Vertices $v(1), \ldots, v(k)$ of a tree define a spanning (sub)tree. Take each maximal path $(w_0, w_1, \ldots, w_\ell)$ in the spanning tree whose intermediate vertices have degree 2, and contract to a single edge of length ℓ. Applying this to k independent uniform random vertices from a n-vertex tree $\mathcal{T}_n$, then rescaling edge-lengths by the factor $n^{-1/2}$, gives a tree we'll call $\mathcal{S}(n,k)$. We visualize such trees as below, vertex $v(i)$ having been relabeled as i.



Trees are "abstract", not embedded in $\mathbb{R}^2$.

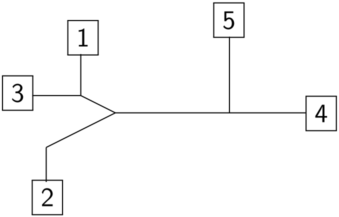


For suitable models of random n-vertex tree $\mathcal{T}_n$, there is a limit

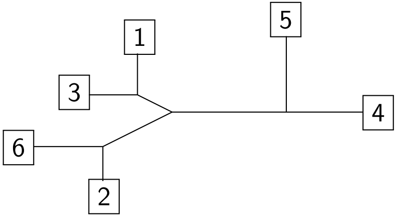
$$\mathcal{S}(n,k) \ \xrightarrow{d}\ \mathcal{S}(k) \text{ as } n \to \infty \text{ with fixed } k.$$

(i) The state space is the space of trees with k leaves labeled $1, 2, \ldots, k$ and with unlabeled degree-3 internal vertices, and where the $2k-3$ edge-lengths are positive real numbers.

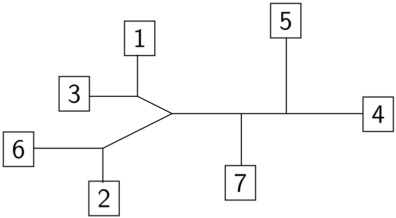
(ii) For each possible topological shape, the chance that the tree has that particular shape and that the vector of edge-lengths $(L_1, \ldots, L_{2k-3})$ is in $([l_i, l_i + dl_i], 1 \leq i \leq 2k-3)$ equals $\quad s \exp(-s^2/2) dl_1 \ldots dl_{2k-3}$, where $s = \sum_i l_i$.



From the "sampling" construction (recall the general program) the distributions of the $\mathcal{S}(k)$ must be consistent as k varies, and indeed there is a simple rule for how to add a new edge to $\mathcal{S}(k)$ to get $\mathcal{S}(k+1)$. Using this rule we can build a random tree with a countable infinite number of leaves $k = 1, 2, 3, \ldots$. Finally take a closure to get what is now called the (Brownian) **continuum random tree** (CRT).



From the "sampling" construction (recall the general program) the distributions of the $\mathcal{S}(k)$ must be consistent as k varies, and indeed there is a simple rule for how to add a new edge to $\mathcal{S}(k)$ to get $\mathcal{S}(k+1)$. Using this rule we can build a random tree with a countable infinite number of leaves $k = 1, 2, 3, \ldots$. Finally take a closure to get what is now called the (Brownian) **continuum random tree** (CRT).



From the "sampling" construction (recall the general program) the distributions of the $\mathcal{S}(k)$ must be consistent as k varies, and indeed there is a simple rule for how to add a new edge to $\mathcal{S}(k)$ to get $\mathcal{S}(k+1)$. Using this rule we can build a random tree with a countable infinite number of leaves $k = 1, 2, 3, \ldots$. Finally take a closure to get what is now called the (Brownian) **continuum random tree** (CRT).

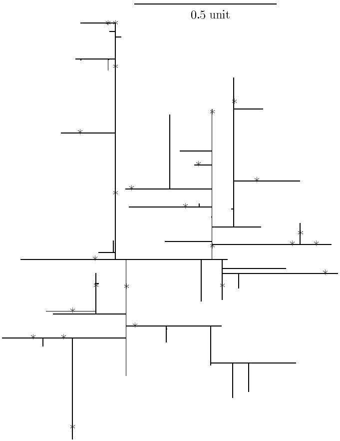


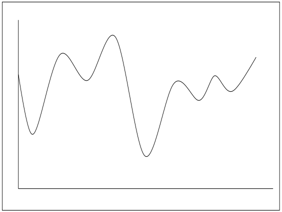
Figure 2

There's a more concrete way to construct such **real (continuum) trees**, observed by Aldous (1991) and Le Gall (1991) .

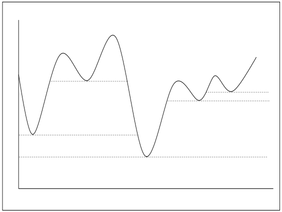
Consider a continuous excursion-type function $f : [0,1] \to [0,\infty)$ with $f(0) = f(1) = 0$ and $f(x) > 0$ elsewhere. Use f to define a continuum tree as follows. Define a pseudo-metric on $[0,1]$ by:

$$d(x,y) = f(x) + f(y) - 2\min(f(u) : x \le u \le y), \qquad x \le y.$$

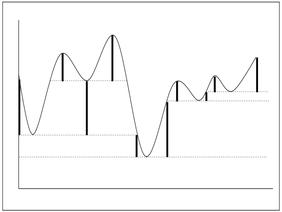
The continuum tree is the associated metric space. Using this construction with standard Brownian excursion (scaled by a factor 2) gives the Brownian CRT.



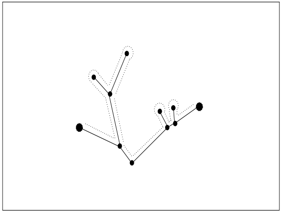
(a)



(b)



(c)



(d)

To summarize: the **Portmanteau Theorem** gives 4 constructions of the CRT.

1. Sequence of spanning subtrees $(\mathcal{S}(k))$ has prescribed law.
2. A certain line-breaking construction gives the sequence in **1**
3. Construction from Brownian excursion and the general function $\leftrightarrow$ tree map.
4. Rescaled weak limit of various families of n-vertex random trees.

Our focus in this talk was on **1** – get a limit object via induced substructures on sampled vertices – but the CRT illustrates general goal of identifying such limit objects with more concrete representations.

3 novel pure math developments

Over 2004-8 there were 3 independent rediscoveries of the basic structure theory, motivated by "pure math" questions in different fields and leading in novel directions.

I'll say (only) a few words about each.

(i) Isometries of metric spaces with probability measures

Question: Can we characterize a "metric space with probability measure" up to measure-preserving isometry? That is, can we tell whether two such spaces (S_1, d_1, μ_1) and (S_2, d_2, μ_2) have a MPI?

The analog is difficult for "metric space" but easy for "metric space with probability measure". Given (S, d, μ), take i.i.d.(μ) random elements $(\xi_i, 1 \leq i < \infty)$ of S, form the array

$$X_{\{i,j\}} = d(\xi_i, \xi_j); \ \{i,j\} \in \mathbb{N}_{(2)}$$

and let Ψ be the distribution of the infinite random array. It is obvious that, for two isometric "metric spaces with probability measure", we get the same Ψ, and the converse is a simple albeit technical consequence of the uniqueness part of structure theory, implying:

Theorem (Vershik (2004))

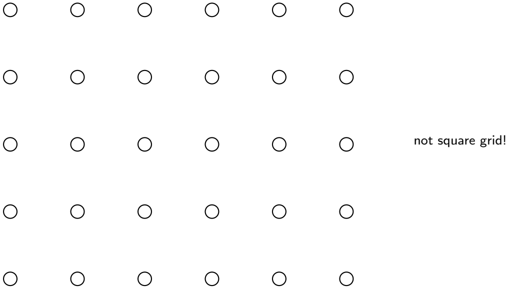
"Metric spaces with probability measure" are characterized up to isometry by the distribution Ψ.

(ii) Limits of dense graphs

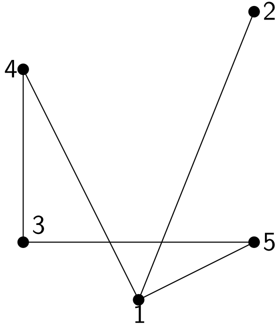
Being a probabilist, I visualize the underlying "size n structures as being **random**, but one can actually apply our "general scheme" to some settings where they are **deterministic**. (Recall we introduce randomness via random sampling).

Here's the simplest interesting case.

Suppose that for each n there is a graph G_n on n vertices.
But we don't see the edges of G_n.
Instead we can sample k random vertices and see the induced subgraph on the sampled vertices.



Induced subgraph $\mathcal{S}(n, k)$ on k of the n vertices of G_n.



One sense of "convergence" of graphs G_n is that for each fixed k the random subgraphs $\mathcal{S}(n, k)$ converge in distribution to some limit $\mathcal{S}(\infty, k)$.

This fits our setup, as follows ......

Suppose that for each n we have a graph G_n on n vertices.
For each n let $(U_{n,i}, i \geq 1)$ be i.i.d. uniform on vertex-set of G_n. Consider the infinite $\{0,1\}$-valued matrix $\mathbf{X}^n$:

$$X^n_{i,j} = 1((U_{n,i}, U_{n,j}) \text{ is an edge of } G_n).$$

When $n \gg k^2$ the k sampled vertices $(U_{n,1}, \ldots, U_{n,k})$ of G_n will be distinct and the $k \times k$ restriction of $\mathbf{X}^n$ is the incidence matrix of the induced subgraph $\mathcal{S}(n,k)$ on these k vertices. Suppose there is a limit random matrix $\mathbf{X}$:

$$\mathbf{X}^n \ \stackrel{d}{\to} \ \mathbf{X} \text{ as } n \to \infty \tag{4}$$

in the usual product topology, that is

$$(X^n_{i,j}, \ 1 \leq i,j \leq k) \ \stackrel{d}{\to} \ (X_{i,j}, \ 1 \leq i,j \leq k) \text{ for each } k.$$

- By compactness there is always a *subsequence* in which such convergence holds.
- For a non-trivial limit we need the **dense** case where (number of edges of G_n)$/\binom{n}{2} \to p \in (0,1)$.

Now each $\mathbf{X}^n$ has the partially exchangeable property (2), and the limit $\mathbf{X}$ inherits this property, so we can apply the representation theorem to describe the possible limits. In the $\{0,1\}$-valued case we can simplify the representation. First consider a representing function of form (3) but not depending on the first coordinate – that is, a function $f(u_i, u_j, u_{\{i,j\}})$. Write

$$q(u_i, u_j) = \mathbb{P}(f(u_i, u_j, u_{\{i,j\}}) = 1).$$

The distribution of a $\{0,1\}$-valued partially exchangeable array of the special form $f(U_i, U_j, U_{\{i,j\}})$ is determined by the symmetric function $q(\cdot,\cdot)$, and so for the general form (3) the distribution is specified by a probability distribution over such symmetric functions.

This all fits our "general program". From an arbitrary sequence of finite deterministic graphs we can (via passing to a subsequence if necessary) extract a "limit infinite random graph" $\mathcal{S}_\infty$ on vertices $1, 2, \ldots$, defined by its incidence matrix $\mathbf{X}$ in the limit (4), and we can characterize the possible limits.

What is the relation between $\mathcal{S}_\infty$ and the finite graphs (G_n)? In probability language it's just

> *the restriction $\mathcal{S}_k$ of $\mathcal{S}_\infty$ to vertices $1, \ldots, k$ is distributed as the $n \to \infty$ limit of the induced subgraph of G_n on k random vertices.*

A recent line of work in graph theory, initiated by Lovász - Szegedy (2006), started by defining convergence in a more combinatorial way (counting number of subgraphs of G_n homomorphic to fixed graphs) which is equivalent (see Diaconis-Janson (2008)) to our notion of G_n converging to $\mathcal{S}_\infty$. They rediscovered the structure theorem, and have used it to develop new and interesting results in graph theory.

(iii) Further uses in finitary combinatorics

The remarkable recent survey by Austin (2008) gives a more sophisticated treatment of the theory of representations of jointly exchangeable arrays, with the goal of clarifying connections between that theory and topics involving limits in finitary combinatorics.

In particular, Austin (2008) describes connections with the "hypergraph regularity lemmas" featuring in combinatorial proofs of Szemerédi's Theorem, and with the structure theory within ergodic theory that Furstenberg developed for his proof of Szemerédi's Theorem.

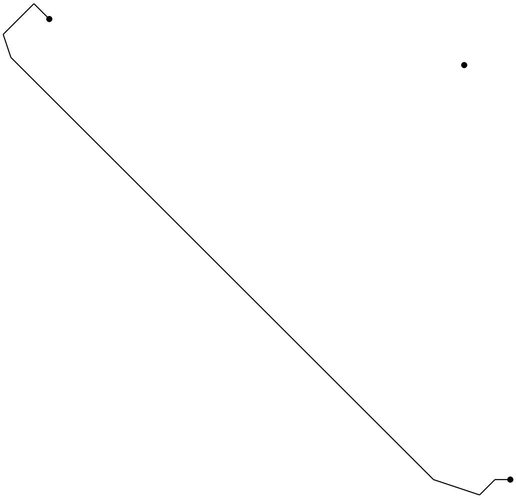
Subsequently Austin - Tao (2010) apply such methods to the topic of hereditary properties of graphs or hypergraphs being testable with one-sided error; informally, this means that if a graph or hypergraph satisfies that property "locally" with sufficiently high probability, then it can be modified into a graph or hypergraph which satisfies that property "globally".

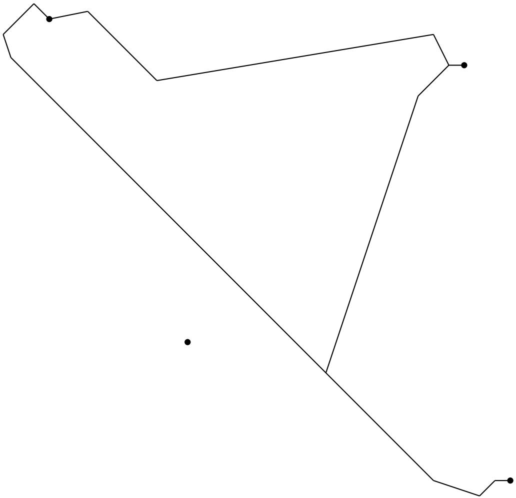
Continuum spatial random networks

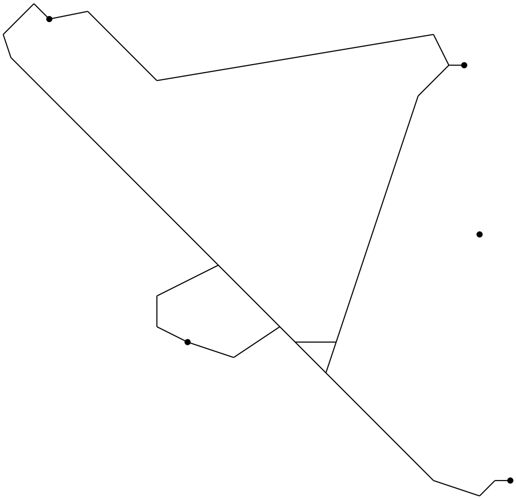
•

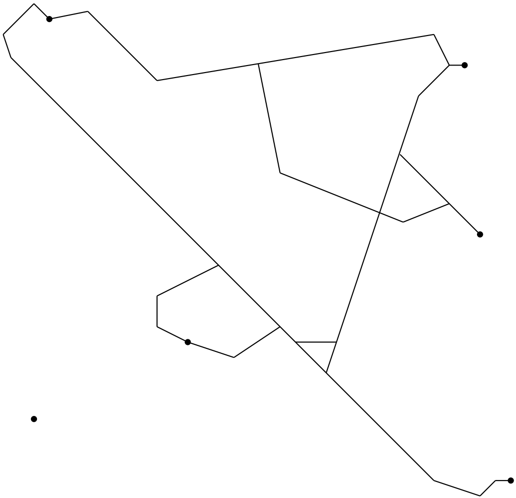
Take two addresses in U.S. and ask e.g. *Google maps* for a route between them.

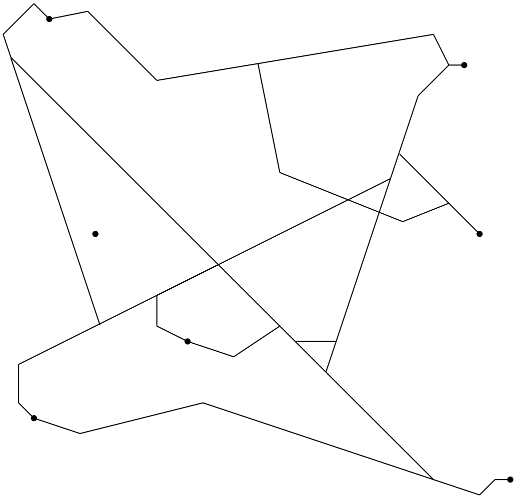
•

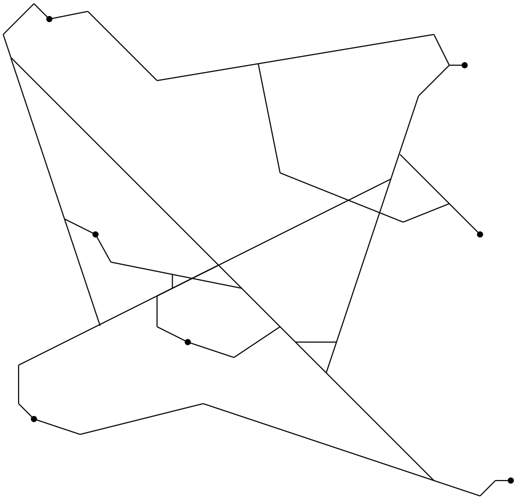












Recall **rather obvious idea:**

> *One way of examining a complex mathematical structure with a PM is to sample IID random points and look at some form of induced substructure relating the random points*

Instead of a real-world road network let's imagine a mathematical model for a random network with routes between (almost) all pairs of points in the plane. We can then pick k random points and look at the induced sub-network.

[Work in progress with Wilfrid Kendall]

Model; for each pair of points (z, z') in the plane, there is a random route $\mathcal{R}(z, z') = \mathcal{R}(z', z)$ between z and z'.

The process distribution (FDDs only) has
(i) translation and rotation invariance
(ii) scale invariance.

Scale invariance implies that the route-length D_r between points at distance r apart must scale as $D_r \stackrel{d}{=} rD_1$, where of course $1 \leq D_1 \leq \infty$.

The statistic $\mathbb{E}D_1$ indicates how effective the network is in providing short routes. We are interested in the case

$$\mathbb{E}D_1 < \infty.$$

There are two other statistics we want to be finite.

Question: How can we study "normalized length" for such a network?
Answer: We explore the network via the subnetwork on a Poisson process of points.

Write $\mathcal{S}(\lambda)$ for the subnetwork on a Poisson (rate λ per unit area) point process. Then **scale-invariance** gives a distributional relationship between $\mathcal{S}(\lambda)$ and $\mathcal{S}(1)$.

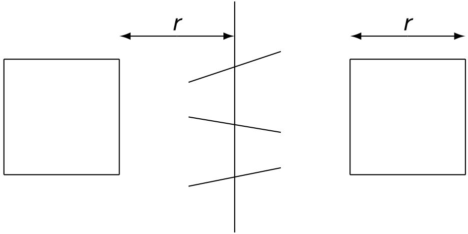
Define normalized length L as mean length-per-unit-area of $\mathcal{S}(1)$.

We are interested in the case

$$L < \infty.$$

Define $p(\lambda, r)$ as length-per-unit-area of segments in $\mathcal{S}(\lambda)$ which are on route between some two points at distance r from the segment.
Set $p(r) := \lim_{\lambda \uparrow \infty} p(\lambda, r)$. Scale-invariance implies $p(r) = p(1)/r$.
We are interested in the case (underlying idea illustrated in picture)

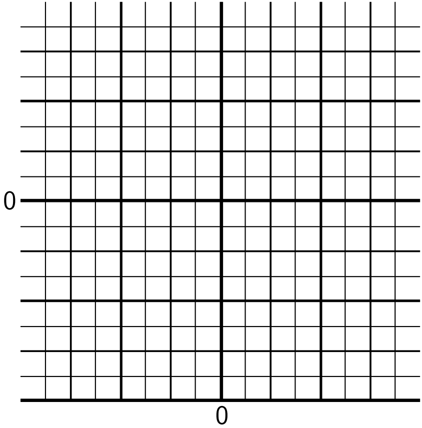
$$p(1) < \infty.$$



Question: do there exist networks with

$$1 < \mathbb{E} D_1 < \infty; \quad L < \infty; \quad p(1) < \infty.$$

Answer: Yes, but we don't know any that is tractable enough to do concrete calculations. I'll outline one construction and mention a second.



Start with square grid of roads, but impose "binary hierarchy of speeds": a road meeting an axis at $(2i+1)2^s$ has speed limit γ^s for a parameter $1 < \gamma < 2$. Use "shortest-time" routes.

(weird – axes have infinite speed limits!)

"Soft" arguments extend this construction to a scale-invariant network on the plane.

- Consistent under binary refinement of lattice, so defines routes between points in $\mathbb{R}^2$.
- Force translation invariance by large-spread random translation.
- Force rotation invariance by randomization.
- Invariant under scaling by 2; scaling randomization gives full scaling invariance.

Need calculations (bounds) to show finiteness of the parameters.

Topic interesting as "symmetry-breaking"; Euclidean-invariant problem on $\mathbb{R}^2$ but any feasible solution must break symmetry to have freeways.